

Company-Wide Data Integration Model for the Water Utility Sector

Oarabile Sebuli, Otlhpile Dinakenyane, Dimane Mpoeleng
Department of Computer Science & Information Systems
Botswana International University of Science & Technology
Palapye, Botswana

Abstract— Data in the water utility sector are generated by different tools of different types in different formats and stored in different systems that run in different platforms and located in different places. This limits the value-adding potential of data thus posing a need for company-wide integrative approaches to data management. Though data warehousing proves to be the best solution in this regard, prior models reviewed in this paper offer limited integrative capabilities with focus on unique operational functions. Therefore, this paper advances a data integration model designed according to Kimball's Business Dimensional Lifecycle methodology and comprising of the processes of water abstraction, production, storage, distribution, consumption, and the revenue and cost functions as well. The paper aims at establishing the integration capabilities of data integration across the value chain with a specific application to Water Utilities Corporation Botswana (WUC).

Keywords— Data Integration; Data Warehousing; Data Model; Data Mart; Dimensional Modeling; Schema Integration; Water Utility; Water Sector; Value Chain; Data Management

I. INTRODUCTION

Data in organizations are produced and used by all business and or operational functions, hence must be allowed to flow freely across the entire ecosystem [1]. Conversely, most organizations have vast amounts of data but find it difficult to access and use it as it is in diverse formats, reside in diverse files and database structures developed by different vendors, and run on diverse platforms [2]. Studies show that due to data disintegration, the water utility sector is limited in deriving optimal value from their data [3], [4], [5], [6]. Thus, this sector needs company-wide data management platforms capable of meeting information needs at all managerial levels and that can be achieved through data integration. Data integration pertains to the exchange of information between systems regardless of whether they are designed to work together and this involves the collection, processing, and clearing of data from different sources to generate a unified vision of those data [7]. Even though data warehousing proves to be the best solution to data disintegration, its adoption and use are limited across enterprises [8]. Above that, data warehousing solutions with focus on unique operational functions proves to be ineffective in resolving problems spread across the value chain. Therefore, this work is motivated by the current data integration needs, coupled with the minimal adoption and application of data warehousing and the minimal existence of formal researches on the application of data warehousing in the water utility sector. The objective of this paper is to explore the integration

potentials that can be leveraged in a data warehouse that models water utility value chain with reference to WUC. The development of the model followed Kimball's Business Dimensional Lifecycle in which data marts were created first [9], modelled one at a time [10] and combined at the end to create a single repository [9]. We present model requirements in the form of value chain processes and corresponding data views. We then integrate the business processes and data views into a single repository model for company-wide data management, viewing and analysis. Finally, we discuss findings on the integration capabilities of the model. The rest of the paper is organized as follows: section II overviews data disintegration in the water utility sector from the global, continental (Africa), and country (Botswana) perspectives. Section III gives insights on the data integration platform advanced in this paper, data warehousing. Section IV discusses the proposed model in terms of methodology, conceptual designs and model evaluation. Section V presents discussions on findings. Finally, section VI draws conclusion on research findings.

II. DATA DISINTEGRATION IN THE WATER UTILITY SECTOR

Data disintegration is one of the major challenges facing Botswana, Africa and the global water utility sector. In the global water utility sector, data are generated and or stored in diverse operational systems spread across the value chain such as Hydraulic Modelling tools, Flow and Pressure meters, Geographical Information Systems (GIS) and Schematic tools, Acoustic sensors, Water Quality sensors, Pump Optimization tools, Data Loggers, Supervisory Control and Data Acquisition (SCADA) systems, Asset Management tools, Automatic Meter Reading (AMR) and Advanced Metering Infrastructure (AMI) tools, Leak Detection software, Alert Systems and Workforce tools [11]. The water utility sector also make use of support systems such as Human Resource Management systems, Finance Management systems, Accounting Packages and other back office applications that support daily operations [12]. These operational systems operate as silos, thus making it difficult to summarize activities across business processes [10]. For instance, the strategic approach to data management and modelling of raw surface water and drinking water systems differ significantly as they are considered distinct operational functions, thus making it difficult to integrate them for comprehensive watershed decisions [13]. Also, although water quantity and quality are inseparable, they are managed by different authorities, thus creating numerous obstacles between

water users and managers [14]. In addition, this sector uses Enterprise Applications such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) and Supply Chain Management (SCM) systems [15] that integrate an organization's business processes with those of its suppliers and customers [16]. Though Enterprise Applications achieve data integration, they are optimized for data input and transaction processing rather than for reporting and analytics [15], hence susceptible to the weaknesses of operational systems. Moreover, the level of disintegration increases as data grow 'old' and are archived for operational efficiency [17]. The global water utility sector has come to realize that data integration is crucial to its operations and sustenance and this has given birth to global initiatives such as the United Nations' Integrated Water Resources Management (IWRM) [18]. In its deliberation on the current state of affairs, IWRM initiative poses a need for processes' integration across the entire water utility value chain:

The traditional fragmented approach is no longer viable and a more holistic approach to water management is essential. This is the rationale for the Integrated Water Resources Management (IWRM) approach that has now been accepted internationally as the way forward for efficient, equitable and sustainable development and management of the world's limited water resources ... [4].

The continent of Africa is hard-hit by the global data disintegration crisis. For instance, the Southern African Development Committee (SADC) reports challenges such as incompatible water resources data and information systems, limited coverage and poor operation and maintenance of hydro-meteorological networks between riparian states. Also, sharing information and hydro-meteorological data between shared watercourse states is constrained by technological diversity among member states, diversity in terms of data and information systems, and inaccessibility of relevant data and information to stakeholders [6]. In particular, one of the main stakeholders in the provision of water services in Botswana (SADC member state), WUC, generates, stores and manages a wide diversity of data currently distributed across various manual and electronic systems, formats, platforms and physical places. It is running on SAP ERP system for most of its operational needs: Billing, Human Resource Management (HRM), Material Management, Plant Maintenance, Customer Relationship Management (CRM), Financial Management, and Sales and Distribution. Instrumentation data and data on water abstraction, production, storage and distribution is extracted from source systems (SCADA, in most cases) into spreadsheets for analysis and reporting. While data on production quantity is generated mostly by telemetry systems, production quality data is produced manually as a product of laboratory tests and stored in a repository different and separate from that which holds quantity data. Records of incidents such as pipe bursts are recorded in SAP ERP system but most of the service data are kept in paper forms. Water consumption data are collected on monthly basis, in most cases manually and in some few cases, electronically, through the use of Automatic Meter Reading tools. Consumption data are stored in the Meter Reading Management System, RouteMaster, which is built on Microsoft.Net technology and runs on the Microsoft SQL

Server platform [19]. Routemaster is in turn interfaced with the Billing module which runs on the SAP platform. Some data such as hydro-meteorological, socio-economic, and demographic data are with external stakeholders and service providers, and in social media platforms such as Facebook. Moreover, WUC continuously archives 'old' data, meaning that most of its historical data remain inactive. With this diversity of disintegrated data, it becomes difficult to have a holistic view of value chain operations. It is also difficult to maximize value from the data asset. Therefore, a unifying approach is needed in this sector to efficiently manage, validate and present past and current data collected both internally and externally from various devices in useful formats. This would improve data access and usage and also optimize on the value-adding potentials of data in this sector.

III. WHY DATA WAREHOUSING?

Data warehousing offers tremendous data integration capabilities. This process involves data extraction from internal and external source systems that represent diverse data structures and technical platforms [20]. Thereafter, it undergoes transformations and then loaded into a single repository for ease of understanding, access, and use [2], [20]. Data warehousing separates Online Analytical Processing (OLAP) functions from Online Transactional Processing (OLTP) functions [17] for optimal operational and analytic performance. It also provides a platform for easy integration and maintenance of current and historical data [17]. Despite these potentials, previous reviewed models offer limited capabilities due to specific focus on unique operational functions. The Intelligent Asset Management system enables effective management of asset criticality, asset identification and hierarchy, cost assignment, general ledger reconciliation, and reliability engineering [21]. On the basis of consumption, revenue and cost data, the system built on a data warehouse by Valor Waters (VW) initially addressed financial issues of conservation, revenue risks and affordability. Valor Waters analytic capabilities have evolved with time to support new service offerings such as water cut off analyzer, handles water cut-offs for non-paying customers; water rate simulator, simulates water usage; hidden revenue locator, eliminates water meter discrepancies; water energy nexus calculator, calculates water energy usage, and hot water heater leak detector which detects water heater leaks [22]. Data warehousing has also been applied to help water utility organizations maintain their immense water infrastructure system by aiding in information analysis about water works systems and determining the best ways to maximize maintenance budget [23]. There is also an innovation where a data warehouse serves as the foundation for District Meter Area (DMA) flow meter data analysis system for detection of bursts at the DMA level and generation of automated alerts in response to abnormal flows [24]. To address the integration limitations of process-specific data models, this paper proposes data warehousing across the water utility value chain for company-wide integration platform.

IV. THE PROPOSED DATA INTEGRATION MODEL

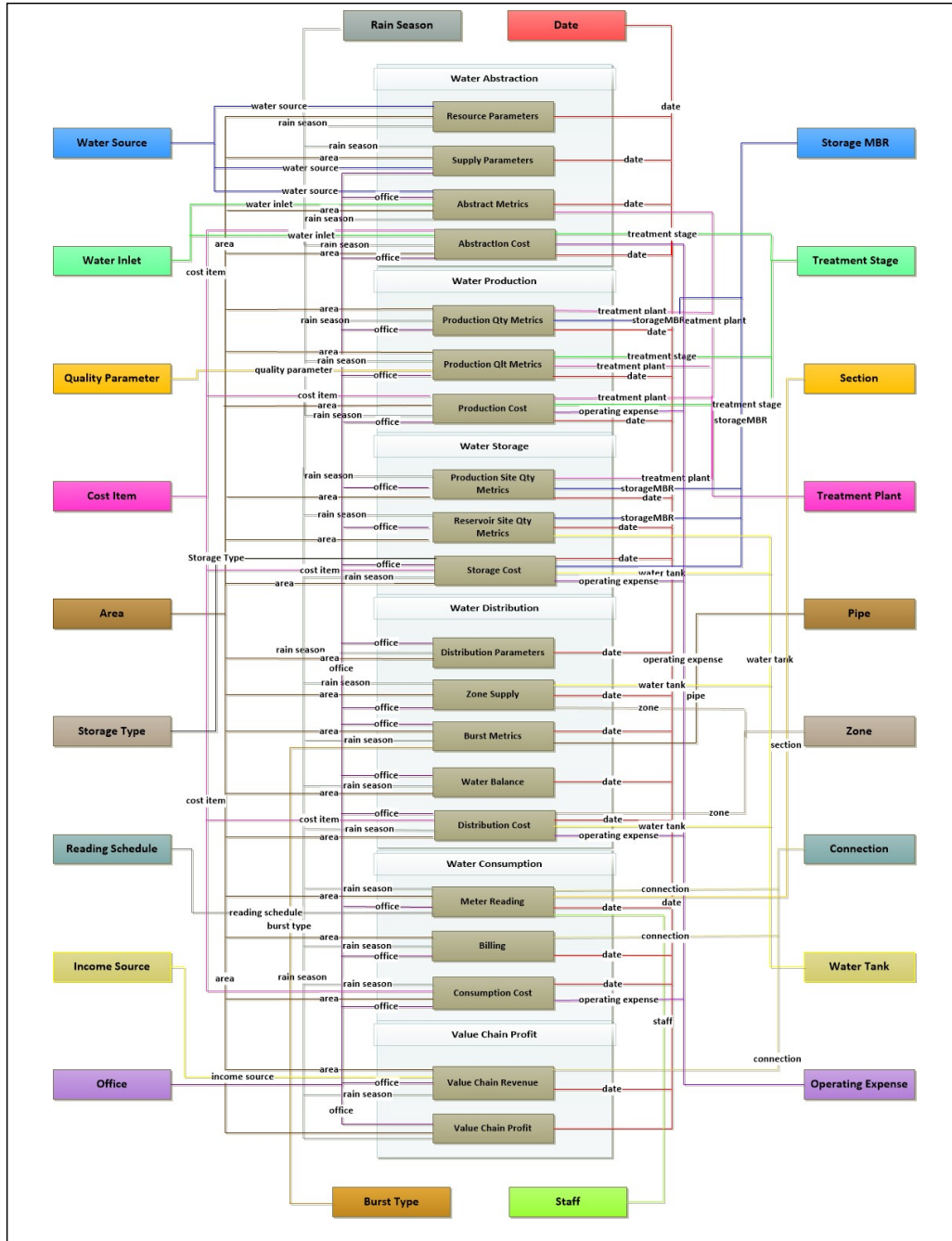


Figure 1: Proposed data warehouse model for WUC

A. Methodology

The proposed data warehousing model was constructed on the basis of the water utility value chain processes of water abstraction, production, storage, distribution and consumption. Kimball's Business Dimensional Lifecycle methodology was adopted in the design of the model. The first activity in this approach is the identification of information requirements and associated business processes. The deliverable at the end of this activity is a document called Data Warehouse Bus Matrix (see figure 2), listing key business processes and some insights on how they are to be analyzed. Then data marts, data models representing individual operational functions, are built based

B. Model Designs

The first data mart, water abstraction integrates water abstraction quantity metrics with water resource parameter metrics (i.e. dam level), supply parameter metrics (contractual supply volumes) and water abstraction costs. The second data mart, water production integrates production quantity with quality and costs. The third data mart, water storage unifies production site metrics with distribution/reservoir site metrics and storage costs. The fourth data mart, water distribution combines distribution parameters with the different distribution functions (zone supply, pipe bursts, water balance), and distribution costs. The fifth data mart, consumption integrates

Data Warehouse Bus Architecture																				
Water Utility Data Warehouse Value Chain Matrix																				
Value Chain Process	COMMON DIMENSIONS																			
	date	rain season	receiving site/role	income source	area	zone	office	action	staff	operating expense	connection	water source	water inlet	water tank	pipe	treatment plant	treatment stage	storage MBR	storage type	quality parameter
water abstraction	X	X		X			X		X		X	X				X	X			X
water production	X			X			X		X							X	X	X	X	X
water storage	X	X					X		X				X					X	X	X
water distribution	X	X		X	X	X			X				X	X	X					X
water consumption	X	X	X				X	X	X	X										X
value chain profit	X	X		X	X		X			X										

Figure 2: Proposed Bus Matrix for WUC

on the matrix, one at a time and integrated with the addition of each data mart, with the last one completing the development of the data warehouse [9]. This approach uses dimensionality modelling; a logical design technique that maps aspects of business/operational processes [25] and presents data in a standard, intuitive form for high-performance access [9]. In this research, information requirements and associated business processes were identified through literature reviews, unstructured interviews and observations, and deduced for requirements specification, outlined in the data warehouse bus matrix in figure 2.

meter reading metrics with billing and consumption costs while the sixth data mart, value chain profit integrates revenues and costs across the value chain. Finally, all data marts are integrated into a single repository illustrated in figure 1 which demonstrates logical relationships that spans through the entire value chain. Relationship links represent data marts' consolidation through shared dimensions and this involves defining a similar level of granularity for linking together schemas per data mart and the related measures and dimensions. Integrating abstraction metrics with production, storage, consumption, distribution and profitability metrics creates a single view of the value chain represented by the cube

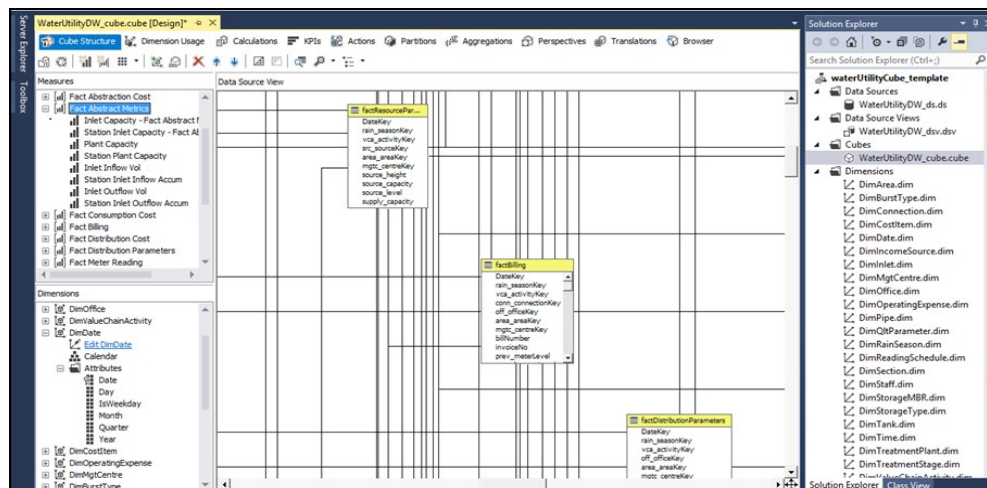


Figure 3: Proposed Integration Cube for WUC

in figure 3 in which measures and dimensions across the value chain are contained in one cube. This data architecture enables water performance to be viewed holistically across the value chain.

C. Model Evaluation

The model was evaluated according to the software testing process encompassing process verification and validation. Verification determines that a system meets the specified requirements while validation evaluates that a system attains its intended use when placed in the intended environment [26]. In process verification, the logical designs were interpreted to WUC participants for evaluation of their correctness in representing water utility value chain and necessary adjustments were made. In process validation, a test case in the form of multidimensional query was developed and run on the integration cube in figure 3. This test case which represents monthly abstraction, production, storage, distribution and consumption costs in a given rain season, screened by cost

enables tracking of relationships between water abstraction metrics, supply parameter metrics, water resource parameter metrics and resultant costs at a diversity of 10 analytic views. Water production integration model provides a platform for tracking and analyzing production quantity in relation to quality and costs at a dynamicity of 10 data views also. Water storage integration enables the flow of water to be tracked between the production site and distribution/reservoir site as well as the resulting costs at 10 multi-angle views. With water distribution integration, zone supply metrics may be viewed in relation to the set distribution parameters, pipe bursts and distribution costs metrics at a diversity of 11 analytic angles. Integrating the water consumption function enables meter reading, billing and resulting costs to be analyzed at a diversity of 10 analytic angles. Finally, the integration of value chain revenues and costs provides a platform for profitability analysis at 6 various angles. In overall, the data warehouse model represents a company-wide integration model with 22

Filtering Dimensions											
RainSeason	All	+									
MgtCentreByRegion	All	+									
AreaByCluster	All	+									
OfficeByType	All	+									
Row Dimensions											
	AbstractExpenseQty	AbstractExpenseAmnt	PdctExpenseQty	PdctExpenseAmnt	StorageExpenseQty	StorageExpenseAmnt	DistribExpenseQty	DistribExpenseAmnt	ConsumptExpenseQty	ConsumptExpenseAmnt	
2015	52670545920	22690488320	1035126411	38247907328	33579298816	21353484288	94128291840	34610921472	53310554112	22129786880	
January	12856344576	6237569024	26103242752	10569642688	7872053248	4738588672	25144688640	9834938368	12662013952	5889212416	
Treatment & Distribution	4222205952	2092843008	8395931648	3579658240	2741059584	2184323072	7564312576	3267133440	3801709440	1938669568	
Electricity Use	4222107648	56288704	209882912	894914560	685264896	546080768	1891078144	816783360	950273960	484667392	
production	4222107648	56288704	209882912	894914560	685264896	546080768	1891078144	816783360	950273960	484667392	
Power	4222107648	56288704	209882912	894914560	685264896	546080768	1891078144	816783360	950273960	484667392	
Labor Use	98304	1529954304	6296948736	2684743680	2055794688	1638242304	5673234432	2450350080	2852782080	1454002176	
staff	98304	1529954304	6296948736	2684743680	2055794688	1638242304	5673234432	2450350080	2852782080	1454002176	
Salaries & Wages ind.	81920	973291520	4197965824	1789829120	1370527972	1092161536	3782156288	1633566720	1901854720	969334784	
Salaries & Wages P&P	16384	55662784	209882912	894914560	685264896	546080768	1891078144	816783360	950273960	484667392	
February	4572987392	2110816256	9153488848	3575300096	2539051512	1159405568	9033416704	3281618896	4511760384	2043117568	
Treatment & Distribution	4572987392	2110816256	9153488848	3575300096	2539051512	1159405568	9033416704	3281618896	4511760384	2043117568	
Electricity Use	4572889088	548831232	2288371712	893825024	638476288	289851392	2258354176	820404224	1127940096	510779392	
production	4572889088	548831232	2288371712	893825024	638476288	289851392	2258354176	820404224	1127940096	510779392	
Power	4572889088	548831232	2288371712	893825024	638476288	289851392	2258354176	820404224	1127940096	510779392	
Labor Use	98304	1561985024	6865115136	2681475072	1915428864	869554176	6775062528	2461212672	3383820288	1532338176	
staff	98304	1561985024	6865115136	2681475072	1915428864	869554176	6775062528	2461212672	3383820288	1532338176	
Salaries & Wages ind.	81920	1005322240	4576743424	1787650048	1276952576	579702784	4516708352	1640808448	2255880192	1021558784	
Salaries & Wages P&P	16384	55662784	2288371712	893825024	638476288	289851392	2258354176	820404224	1127940096	510779392	
March	4061151232	2033909760	853824256	3414884352	2577088512	1394800352	8546959360	3286188032	4344544128	1907425280	
Treatment & Distribution	4061151232	2033909760	853824256	3414884352	2577088512	1394800352	8546959360	3286188032	4344544128	1907425280	
Electricity Use	4061052928	571211776	2138456064	853721088	644272128	348715008	2136739840	821547008	1086636032	476856320	
production	4061052928	571211776	2138456064	853721088	644272128	348715008	2136739840	821547008	1086636032	476856320	
Power	4061052928	571211776	2138456064	853721088	644272128	348715008	2136739840	821547008	1086636032	476856320	
Labor Use	98304	1462697984	6415368192	2561163264	1932816384	1046145024	6410219520	2464641024	325908096	1430568960	
staff	98304	1462697984	6415368192	2561163264	1932816384	1046145024	6410219520	2464641024	325908096	1430568960	
Salaries & Wages ind.	81920	989593600	4276912228	1707442176	128954256	697430016	4273479680	1643094016	2173272064	953712640	
Salaries & Wages P&P	16384	473104384	2138456064	853721088	644272128	348715008	2136739840	821547008	1086636032	476856320	

Figure 4: Integration Test Case Illustration

item, operating expense, office (customer service center/water works station/reservoir site), area, and management center is illustrated in figure 4.

V. FINDINGS AND DISCUSSIONS

We found out that the proposed model provides a platform for maximal integration of data at individual operational functions (represented by data mart models) and at the corporate level (represented by the data warehouse model). Each data mart has one or more schemas (individual schemas represent integration of individual sub-functional operations within an operational function) connected through shared dimensions for holistic integration of operational processes modelled by the data mart. Water abstraction integration model

dimensional views. For instance, figure 4 demonstrates costs across the value chain under a diversity of data views. Water quantity may also be traced from the beginning process of water abstraction to the end process of consumption, thus giving a complete picture of what happens to water as it traverses the network. Tracing water volumes from the beginning to the end of the value chain may aid in addressing the issue of water loss/non-revenue water. Water loss may be tracked along the value chain to identify processes with high or low water losses in relation to diverse dimensional attributes. The differences in those losses may also be accounted for in terms of the dimensional variations represented by the model. Tracing water quantity flows along the value chain may also aid in computing the water balance statement to account for

water usage given the system input volume at a given time and other dimensional variations. Moreover, this model enables water quality to be traced alongside water quantity from water inlet (pre-treatment stage) through the different water processing stages and distribution networks to identify specific places of high or low infection risks.

VI. CONCLUSION

The data warehouse that models the value chain has the capability to turn the water utility sector aspirations of water management efficiency from being an abstract concept to reality. This model provides a platform for optimal data integration at both the data mart and warehouse levels to create a single version of truth that spans the water utility value chain. This model enables data in social media platforms and that with external stakeholders and service providers to be combined with internal data in operational systems, archival storages and physical files, thus, ensuring that all data across the value chain remain active and contribute its value to the information needs of the organization. This process-based data model guarantees stability in dynamism as it is more adaptive to constant adjustments to information needs compared to a data warehouse model designed to address specific information needs at a specific time. The unstable information needs may then be plugged into the model which offers flexibility to easily adjust to the constantly changing information requirements without having to rebuild the data warehouse structure or with minor alterations to existing architecture. Therefore, the model provides a better foundation for company-wide implementations such as balanced score cards. However, for a complete view of water utility operational performance, waste water processes also need to be integrated into the model.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science and Information Systems and Mr. Kelebonye Bagai of BIUST for their valued contribution to this work. The authors are also grateful to Water Utilities Corporation Botswana (WUC) for their invaluable support in opening up their waterworks for process observations and for providing metrics and available data on those processes as well.

REFERENCES

- [1] H. Heyns, "Becoming an analytics-driven organisation to create value," Ernst & Young, United Kingdom, Technical document 2015.
- [2] C. Ballard et al., "Data modeling techniques for data warehousing," IBM, California, Technical documentation 1998. [Online]. <http://www.redbooks.ibm.com/redbooks/pdfs/sg242238.pdf>
- [3] J. S. Horsburgh, D. G. Tarboton, D. R. Maidment, and I. Zaslavsky, "A relational model for environmental and water resources data," *Water Resour. Res.*, vol. 44, no. W05406, pp. 1-12, May 2008.
- [4] United Nations. (2014, November) Integrated water resource management. [Online]. <http://www.un.org/waterforlifedecade/iwrm.shtml>
- [5] Accenture Consulting, "Unlocking the value of analytics: accenture's digitally enabled grid program," Accenture Consulting, Technical document 2013. [Online]. <https://www.accenture.com/us-en/insight-digitally-enabled-grid-unlocking-value-analytics.aspx>
- [6] SADC, "sadc regional water strategy," SADC, Gaborone, Strategy document 2006.
- [7] A. H. Mousa, N. Shiratuddin, and M. S. A. Bakar, "Generic framework for better choosing between data integration types (GFCBDIT) during build business intelligence applications," *International Journal of Digital Content Technology and its Applications (JDCTA)*, vol. 8, no. 5, pp. 27-35, October 2014.
- [8] J. McKendrick. (2011, September) A new dimension to data warehousing: 2011 ioug data warehousing survey. [Online]. <http://www.dbta.com/research>
- [9] T. M. Connolly and C. E. Begg, "Business intelligence," in *Database systems: a practical approach to design, implementation, and management*, 5th ed. London, England: Pearson Education, 2010, ch. 32-35, pp. 1145-1229.
- [10] M. V. Mannino, *Data warehouse technology and management*, 2nd ed. New York, America: McGraw-Hill, 2004.
- [11] S. Callahan and J. M. Miller, "Data integration solutions for effective utility management: progression towards smart city implementation," in *One Water Technical Conference. & Expo*, Ohio WEA-AWWA, 2014.
- [12] P. Ponniah, *Data warehousing fundamentals: a comprehensive guide for IT professionals*. New York: John Wiley & Sons, Inc, 2001.
- [13] S. C. Karcher, J. M. VanBriesen, C. T. Nietch, and M. T. Heberling, "Assessing the challenges associated with developing an integrated modeling approach for predicting and managing water quality and quantity from the watershed through the drinking water treatment system," EPA, Cincinnati, Technical document 2012.
- [14] T. Dai and J. Labadie, "River basin network model for integrated water quantity/quality management," *Journal of Water Resources Planning and Management*, vol. 127, no. 5, pp. 295-305, October 2001.
- [15] R. Sherman, "Don't stop at how, learn why," *DM Review*, p. 26, June 2008.
- [16] J. L. Whitten, L. D. Bentley, and K. C. Dittman, "Enterprise applications," in *Systems analysis and design methods*. New York: McGraw-Hill, 2004, pp. 32-35.
- [17] J. M. A. Navamani, A. Kannammal, and P. R. J. Thangaiyah, "Building multi-dimensional cube to analyze public examination results: a business intelligence approach," *Applied Mechanics and Materials*, vol. 622, no. 11, pp. 11-22, August 2014.
- [18] Department of Water Affairs- Ministry of Minerals, Energy & Water Resources, "Botswana integrated water resources management & water efficiency plan," Botswana Government, Gaborone, Strategy document 2013.
- [19] Consolidated African Technologies, "rma.net - meter management simplified," Consolidated African Technologies, Randburg, Technical document 2016. [Online]. <http://www.cats.co.za/resources/RMA.Net%20Brochure.pdf>
- [20] H. J. Wixom and B. H. Watson, "The current state of business intelligence," *IT Systems Perspectives*, pp. 96-99, September 2007.
- [21] R. V. Buskirk, "Performance management best practises: data management practises for intelligent asset management in a public water utility," September 2011. [Online]. <http://www.we-inc.com>
- [22] Valor Water. (2016) Valor Water website. [Online]. <http://www.valorwater.com>
- [23] Informatica. (2016) Informatica website. [Online]. <http://www.informatica.com/us/customers/thames-water-utilities-ltd.aspx>
- [24] S. R. Mounce and J. B. Boxall, "Implementation of an on-line artificial intelligence district meter area flow meter data analysis system for abnormality detection: a case study," 2010.
- [25] IBM Corporation. IBM Knowledge Centre. [Online]. http://pic.dhe.ibm.com/infocenter/dataarch/v9r1/index.jsp?topic=%2Fco m.ibm.datatools dimensional.ui.doc%2Ftopics%2Fc_dm_design_cycle_1_idbusprocres.html
- [26] Pan, J. (1999) Carnegie Mellon University. [Online]. https://users.ece.cmu.edu/~koopman/des_s99/sw_testing/